



Cornell University
Cornell Institute for
Social and Economic Research

Research Data Centre (FDZ)
of the German Federal
Employment Agency (BA)
at the Institute for
Employment Research (IAB)



IAB@Cornell

Connecting Researchers with Data: Discovery, Documentation, Access and Security

Cornell Institute for Social and Economic Research (CISER);
German Institute for Employment Research (IAB);
and the Cornell Labor Dynamics Institute (LDI)

Researchers Searching for Data

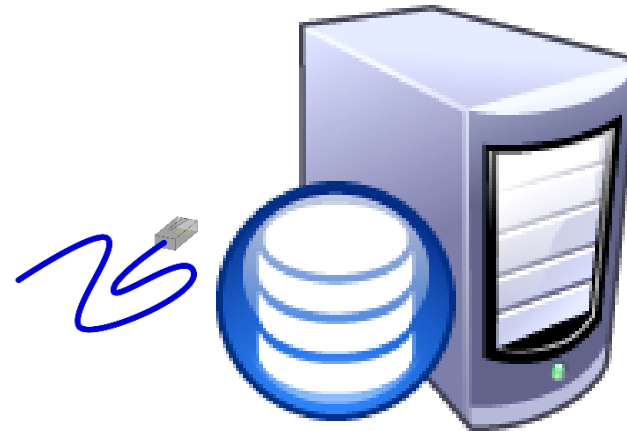


Connecting Researchers with Data

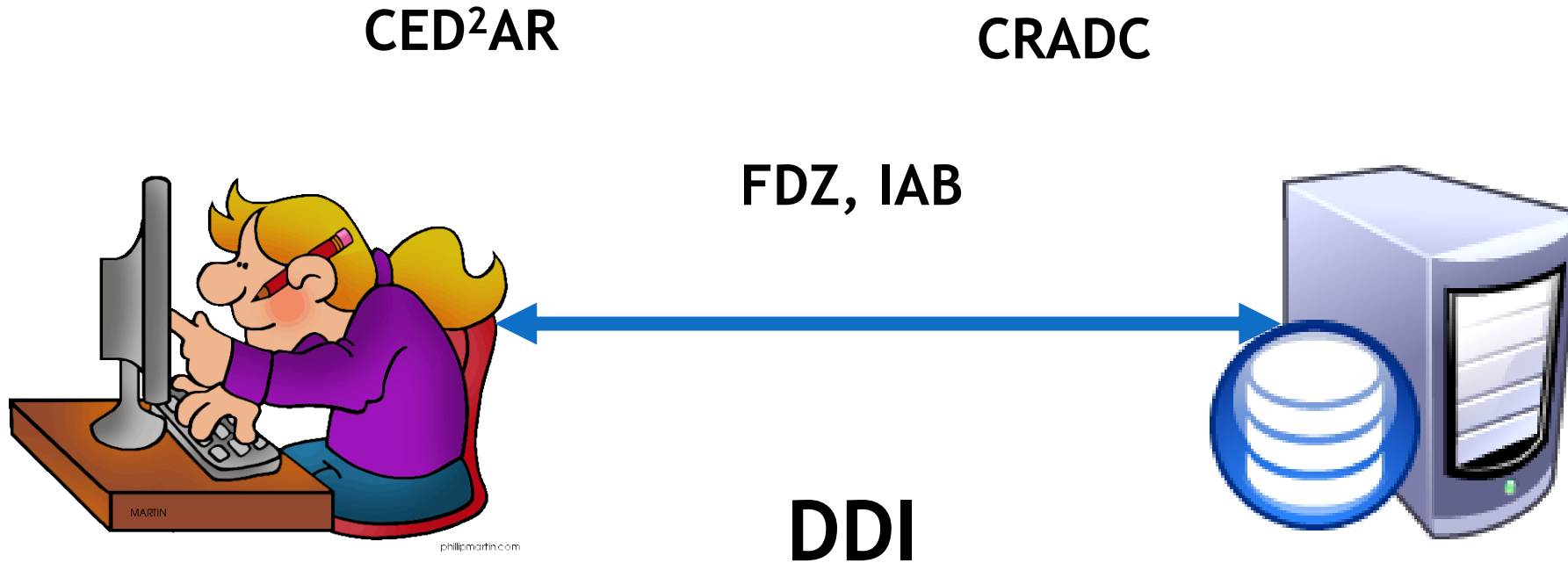
CED²AR

CRADC

FDZ, IAB



Connecting Researchers with Data



Discovering Data File - Do Data Exist??

First barrier to research:

- ▶ A research question is formulated, but are there existing data files to address the question?
- ▶ Discovering data files by searching descriptions of studies, files, and variables

The first barrier is discovery

Restricted Data - What's in that Dataset??

Second barrier to research with restricted data:

- ▶ Researchers need to examine the metadata of a dataset in order to prepare a research plan for the data application
- ▶ Data Providers want research plans to specify why restricted data files are necessary, and how the variables will be analyzed
- ▶ Enclaves and libraries want to assist researchers to search and discover what relevant data files exist in greater detail than abstracts

The second barrier of restricted data research is to discover what's in the file!

Searching Restricted Labor Market Data

- ▶ Labor market datasets for longitudinal employment data
 - ▶ National Longitudinal Surveys (NLSY)
 - ▶ Longitudinal Employer-Household Dynamics (LEHD)
 - ▶ Sample of Integrated Labor Market Biographies (SIAB)
- ▶ That was the study level, but how do we compare the variables?
 - ▶ Searching across variables of restricted data is painful
 - ▶ Codebooks
 - ▶ Abstracts
 - ▶ Having seen the dataset during a prior project
 - ▶ The majority of these resources are not machine readable

There needs to be an accessible discovery tool

CED2AR

- ▶ Comprehensive Extensible Data Documentation and Access Repository
- ▶ Developed by the Cornell Node of the NSF Census Research Network (NCRN)
- ▶ Enables search and browsing across codebooks
- ▶ Designed to improve the discoverability of both public and restricted data
- ▶ Lightweight, DDI driven web application
 - ▶ Based upon leading metadata standards
 - ▶ Ingests data from a variety of sources
- ▶ Study levels include the
 - ▶ Abstract
 - ▶ Variables
 - ▶ Values
 - ▶ Citation
 - ▶ Terms of Use, as applicable

Utilizing CED2AR to Locate SIAB Variables

*The addition of IAB SUF SIAB metadata to CED2AR will enable researchers to search across multiple labor market data series**

- ▶ CISER is currently collaborating with IAB to develop the DDI necessary to add the SIAB metadata to CED2AR
 - ▶ Promote use of the SIAB data through discoverable documentation
 - ▶ Facilitate ease in research plan creation

*<https://www2.ncrn.cornell.edu/ced2ar-web/about>

Filter Codebooks

ACS PUMS ☐ASM ☐BLS ☐BRDIS ☐CBO ☐CMF ☐CPR ☐CRT ☐CSR ☐ECF ☐ECF T26 ☐EDF ☐EHF ☐GAL ☐GAL T26 ☐ICF ☐ICF T26 ☐ILBD ☐LBD ☐

Search

Searching all codebooks. No filters active.

Search

[Advanced Search](#)Show variables

[CED2AR](#) / SIAB-R 7510 SUF Scientific Use File - Sample of Integrated Labour Market Biographies - Regional File 1975-2010

SIAB-R 7510 SUF Scientific Use File

- Sample of Integrated Labour Market Biographies - Regional File 1975-2010

[View Variables](#) (23 variables)

Last update to metadata: 2015-11-18 11:12:39 (auto-generated)

Codebook prepared by: Philipp vom Berge, Anja Burghardt, Simon Trenkle (Institute for Employment Research, IAB)

Data prepared by: Institute of Employment Research, IAB

Data Distributed by:

Institute for Employment Research, IAB

<http://fdz.iab.de>

Citation

Please cite this codebook as:

Institute for Employment Research, IAB. Philipp vom Berge, Anja Burghardt, Simon Trenkle Documentation of labour market data. Sample of Integrated Labour Market Biographies Regional File 1975-2010 (SIAB-R 7510), Version 1 (SIAB_Regional_7510_EN_v1_dok1), September 2013.

Please cite this dataset as:

Sample of Integrated Labour Market Biographies Regional File 1975-2010 (SIAB-R 7510), Nuremberg 2013

Abstract

This FDZ data report describes the "Regionalfile" of the Sample of Integrated Labour Market Biographies 1975-2010 (SIAB-R 7510). This dataset represents the factual anonymous ver-sion of the Sample of Integrated Labour Market Biographies (SIAB) and may be delivered to scientific research institutions after concluding a use agreement with the IAB.

[Search Variables](#)[Browse Variables ▾](#)[Browse by Codebook](#)[Documentation](#)[About](#)

Filter Codebooks

ACS PUMS ☐ASM ☐BLS ☐BRDIS ☐CBO ☐CMF ☐CPR ☐CRT ☐CSR ☐ECF ☐ECF T26 ☐EDF ☐EHF ☐GAL ☐GAL T26 ☐ICF ☐ICF T26 ☐ILBD ☐IBD ☐

Browse all Variables

Searching [SIAB \(v1\)](#) ✕[View in reading mode](#)Show variables

	Variable Name	Label	Codebook
<	<input type="checkbox"/> ao_region	Place of work: district / region	SIAB-R 7510 SUF Scientific Use File - Sample of Integrated Labour Market Biographies - Regional File 1975-2010
	<input type="checkbox"/> begepi	Start date of split episode	SIAB-R 7510 SUF Scientific Use File - Sample of Integrated Labour Market Biographies - Regional File 1975-2010
	<input type="checkbox"/> begorig	Original start date of observation	SIAB-R 7510 SUF Scientific Use File - Sample of Integrated Labour Market Biographies - Regional File 1975-2010
	<input type="checkbox"/> beruf_gr	Occupation - activity performed / last activity aggregated	SIAB-R 7510 SUF Scientific Use File - Sample of Integrated Labour Market Biographies - Regional File 1975-2010
	<input type="checkbox"/> bild	School education and vocational training	SIAB-R 7510 SUF Scientific Use File - Sample of Integrated Labour Market Biographies - Regional File 1975-2010
	<input type="checkbox"/> bnn	Establishment counter	SIAB-R 7510 SUF Scientific Use File - Sample of Integrated Labour Market Biographies - Regional File 1975-2010

CED2AR / SIAB / stib_en

Variable Name

stib_en



git



SAS

Statistica

Label

Occupational status and working hours

Codebook

SIAB

Concept

Type

character

▼ Values (10 total)

0	In vocational training
1	Unskilled worker
2	Skilled worker
3	Master craftsman, foreman (blue-collar or white-collar)
4	Employee (excl. employed master craftsmen)
7	Home workers, Freelance home workers
8	Part-time up to half of full-time
9	Part-time more than half of full-time
n	Systematically not available
Z	No (valid) entry

Restricted Data Access

*The big barrier to restricted data access
is balancing researcher access
with data security*

- ▶ It's ideal when data is housed within a controlled environment
- ▶ It's also ideal when researchers have the most flexible, secure access to the data

IAB SUF Access and Data Security

- ▶ IAB Highly Restricted Data
 - ▶ Originally only accessible at Nuremberg, Germany
 - ▶ Implemented a Research Data Center - in - Research Data Center (RDC-in-RDC) approach
 - ▶ Allowed for a controlled remote access environment from the established Nuremberg research data center to comparably secure research data centers using Citrix-thin client technology
- ▶ IAB SUF Data
 - ▶ Designed specifically by IAB for off-site access
 - ▶ Originally sent to authorized researchers in CD format, later by transmission, to be destroyed at the end of the project
 - ▶ These methods provided access, but posed a security threat via individual accountability and system back-ups
 - ▶ Needed a new, more secure off-site access mode - *CRADC*

Cornell Restricted Access Data Center (CRADC)

- ▶ Cornell Institute for Social and Economic Research (CISER)'s remote access virtual data enclave
 - ▶ Controlled remote access environment accessible worldwide
 - ▶ As a secure computing environment with remote access, CRADC exists to:
 - ▶ House and protect restricted research data
 - ▶ Help PIs comply with requirements of Data Providers
 - ▶ Provide a computing platform as flexible as data use agreements permit
 - ▶ Any approved researcher can securely access restricted project files
 - ▶ Restricted data is available to the researcher on the CRADC server
 - ▶ Restricted data is held on a drive with only read and execute permissions
 - ▶ A project work space is provided on a separate drive, with access to analysis applications
 - ▶ No access is given to email, internet, clipboard, printer sharing, or disk mapping, and analysis can only be saved within the project work space

Institute for Employment Research (IAB)

Scientific Use Files (SUF)

- ▶ Contain confidential administrative microdata on labor markets:
 - ▶ Employment
 - ▶ Unemployment benefit receipts
 - ▶ Participation in labor market programs and registered job search
 - ▶ A large number of socio-economic characteristics
- ▶ Similar in origin to the highly restrictive files but have been factually anonymized, enabling wider access
 - ▶ The highly restrictive files can only be accessed through a thin client connection directly to the IAB servers in Nuremberg, Germany
 - ▶ SUF file access through CISER's remote access virtual data enclave (CRADC) through secure computing accounts using Remote Desktop Connection or Terminal Service Client
- ▶ SUF access is available to any researcher approved by IAB worldwide
 - ▶ ***No affiliation with Cornell is required*** - any researcher approved by IAB can remotely access the files

CRADC - SIAB Data

- ▶ Access the IAB SUF data worldwide
 - ▶ CRADC currently supports SIAB researchers in
 - ▶ Canada
 - ▶ Germany
 - ▶ Spain
 - ▶ United Kingdom
 - ▶ United States
- ▶ CRADC provides the controlled environment for each researcher and project based on IAB's terms
 - ▶ IAB no longer needs to send restricted data CDs or transmissions
 - ▶ Researchers are permitted SFTP access from their project work space to their campus office via static IP address
 - ▶ CRADC can confirm that all project data and analysis files are securely destroyed at the end of any IAB data use agreement

Discovering SIAB through CED2AR, Accessing on CRADC

- ▶ Labor Market Datasets for longitudinal employment data
 - ▶ Discover the study and variables using CED2AR
 - ▶ Search and browse across codebooks
 - ▶ Compare public and restricted dataset versions
 - ▶ Select SIAB
 - ▶ Create a SIAB research plan using CED2AR variables and values
 - ▶ Review the terms of use to confirm that the planned analysis will be accepted (the data security plan is already covered)
 - ▶ IAB approves the longitudinal employment data project
 - ▶ Gain access to the SIAB data through CRADC
 - ▶ Complete SIAB analysis, notify IAB of project completion
 - ▶ CRADC will close the restricted project
- ▶ Browse or search CED2AR for new project variables

Our Team

- ▶ IAB

- ▶ Dana Muller
- ▶ David Schiller
- ▶ Joerg Heining

- ▶ CISER

- ▶ Ben Perry
- ▶ Michelle Edwards
- ▶ Stephanie Jacobs
- ▶ Warren Brown

Promoting
Data through
Discoverability

Data Security

Researcher
Access





Thank you for your time
and attention!



Warren Brown
Stephanie Jacobs

cradc@cornell.edu
ciser.cornell.edu



Cornell University